Estimation in discretely observed diffusions killed at a threshold

Enrico Bibbona*
University of Torino, Torino, Italy
enrico.bibbona@unito.it

Susanne Ditlevsen†
University of Copenhagen, Copenhagen, Denmark.
susanne@math.ku.dk

January 20, 2012

Abstract

Parameter estimation in diffusion processes from discrete observations up to a first-hitting time is clearly of practical relevance, but does not seem to have been studied so far. In neuroscience, many models for the membrane potential evolution involve the presence of an upper threshold. Data are modeled as discretely observed diffusions which are killed when the threshold is reached. Statistical inference is often based on the misspecified likelihood ignoring the presence of the threshold causing severe bias, e.g. the bias incurred in the drift parameters of the Ornstein-Uhlenbeck model for biological relevant parameters can be up to 25–100%. We calculate or approximate the likelihood function of the killed process. When estimating from a single trajectory, considerable bias may still be present, and the distribution of the estimates can be heavily skewed and with a huge variance. Parametric bootstrap is effective in correcting the bias. Standard asymptotic results do not apply, but consistency and asymptotic normality may be recovered when multiple trajectories are observed, if the mean first-passage time through the threshold is finite. Numerical examples illustrate the results and an experimental data set of intracellular recordings of the membrane potential of a motoneuron is analyzed.

Keywords: sequential estimation, diffusion processes, first-passage times, stochastic neuron models, model misspecification, bootstrap

*Partially supported by PRIN 2008
†Supported by grants from the Danish Council for Independent Research | Natural Sciences

1 Introduction

In many applications data from a process are sampled sequentially up to the first time it hits an upper boundary, where some macroscopic event occurs, which renders further observations impossible. A prominent example and the one that triggered the authors interest in the problem comes from neuroscience, where the membrane potential of a neuron is believed to be well described by a diffusion process between the times where the neuron fires. The neuron fires when the membrane potential reaches some upper threshold, and thus, discrete observations of the diffusion process are obtained as long as the process has not crossed the threshold. In this context, statistical inference is often based directly on the longitudinal sample, ignoring the presence of the threshold and that the sample size is now a random variable, see Höpfner (2007); Jahn et al. (2011); Lansky et al. (2006, 2010); Picchini et al. (2008). Such model misspecification, together with the unavoidable shortness of the observed paths, causes severe bias in the estimates of drift parameters (Bibbona et al. (2008, 2010) and Section 3 below), providing a strong motivation to find a dedicated statistical method. Another example comes from biomedicine, where a patient in danger of suffering some attack or death, is regularly monitored, e.g. blood pressure or some substance in the blood, but only when it is between some critical limits, whereafter the subject enters a major crisis. The problem also occurs whenever parameter estimation follows some sequential data acquisition, for example a sequential test (Whitehead, 1986) or a clinical trial (Jung and Kim, 2004; Bibbona and Rubba, 2011). Sequential estimation for i.i.d. samples is a well established field of statistics (Ghosh et al., 1997), but to the authors knowledge, for diffusion processes it has only been studied from continuous observations (Novikov, 1972; Sørensen, 1983; Różański, 1989; Liptser and Shiryaev, 2001) and mainly with the aim of finding *optimal* sampling plans.

For a killed process, no more information can be obtained once the threshold is reached, and thus, standard asymptotic results do not apply. In Sørensen (1983), the likelihood function for a randomly stopped diffusion process is calculated when the process is continuously observed and consistency and asymptotic normality of the MLE are discussed for an increasing sequence of stopping times tending to infinity. In Bhat (1961), regular Markov chains with a finite number of states (and one absorbing) is analyzed and asymptotic results are derived when a large number of trajectories is available. We will extend this result further. In Ferebee (1983), an unbiased estimator is found for the drift parameter of a Wiener process with drift (WD) with unit variance scaling observed continuously up to a stopping time.

The role of the threshold in neuronal modeling was first emphasized in Paninski (2006a,b). In Giraudo et al. (2011) the stochastic differential equation for a diffusion process constrained to remain below the threshold was derived.

In this paper we will study maximum likelihood sequential estimation from a diffusion process discretely observed up to a first-hitting time. In Section 2 the likelihood function is introduced. It is intractable for practically all models except the WD, and we point at some approximations for evaluating it when

the sampling frequency is high. Examples are provided for some models which are of practical interest in many fields such as WD, the Ornstein-Uhlenbeck (OU) and the Square Root (SR) processes. Not surprisingly, when only one trajectory is used in the statistical analysis, estimates are still biased, widely variable and skewed. Bootstrap bias correction may sometimes improve the estimates. To get better estimates one needs to enlarge the sample by considering many independent trajectories. Standard Cramer conditions for consistency and asymptotic normality when the number of trajectories used in the estimation goes to infinity are easily stated, but they are impractical to check. In Section 4 we propose some easier equivalent conditions with a nice probabilistic interpretation in terms of an associated regenerative process. An asymptotic scheme is hence pointed out that allow to find good estimates based on the likelihood function. This solution is of practical relevance in neuroscience where numerous trajectories are usually available and estimating from the global sample resolves any issue. In Section 5 simulation studies are carried out to illustrate the theoretical results and in Section 6 an experimental data set of intracellular recordings of the membrane potential of a motoneuron during mechanical stimulation obtained from an isolated carapace-spinal cord preparation from adult turtles is analyzed. Finally, in Section 7 we summarize our findings and discuss some further directions not treated in the paper. An Appendix provides a few numerical details.

2 The likelihood function

Assume that the discrete time Markov process $\{X_i\}_{i\in\mathbb{N}}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P}_\theta)$ is made of observations from a diffusion process X_t at equidistant time points $i\Delta, i = 0, 1, \ldots, n$ for some fixed $\Delta > 0$. Thus, we write $X_i = X_{i\Delta}$. Assume that X_t is the solution to the stochastic differential equation

$$dX_t = \mu(X_t;\theta)\,dt + \sigma(X_t;\theta)\,dW_t; \quad X_0 = x_0 \tag{1}$$

where W is a Wiener process. The drift and diffusion functions $\mu(\cdot)$ and $\sigma(\cdot)$ are assumed known up to the parameter θ, which belongs to the parameter space $\Theta \subseteq \mathbb{R}^p$. They are assumed to be smooth enough to ensure for every θ uniqueness in law of the solution. Let $f_\theta(y, \Delta|x)$ denote the transition density for a given value of θ, i.e. the conditional density under $\mathbb{P}_\theta$ of $X_{t+\Delta}$ given that $X_t = x$. Assume that the interval $(l, r) \subset \mathbb{R}$, for some $-\infty \le l < r \le \infty$, is the smallest one such that

$$\mathbb{P}\bigcap_{t\ge 0}\big(X_t \in (l, r)\big) = 1$$

and thus the state space is $E = (l, r)$.

Introduce an absorbing threshold at level b with $l < x_0 < b < r$, and assume that the initial value X_0 is fixed and equal to $x_0 \in E_b = (l, b)$. To avoid heavy notations we assume the threshold level is constant, but there is no technical obstruction to threat time-varying thresholds too. Since we have

incomplete observations of a continuous process, we cannot in general ensure that the continuous process has not crossed the threshold between observations, even if all observed points are below the threshold. However, the statistical problem is interesting exactly for those applications where the crossing of the threshold causes some macroscopic event that is always observed, and such that no further observations can be obtained. We assume that we are in this setting, thus the interval where the first-hitting time occurs is always observed.

Define the stopping times

$$\begin{aligned} T_b &= \inf\{t > 0 : X_t \geq b\} \\ N &= \inf\{n \in \mathbb{N} : n\Delta \geq T_b\}. \end{aligned}$$

Throughout the paper, we assume that $P(T_b < \infty \,|\, X_0 = x_0) = P(N < \infty \,|\, X_0 = x_0) = 1$. Let X_t^k be the process that coincides with X_t as long as $X_t < b$ and is killed the first time X_t crosses the threshold, whereafter it goes into a *coffin state* C which is artificially added to the state space, $E_b \cup C$. Let X_i^k be its discretization, i.e.,

$$X_i^k \quad = \quad \begin{cases} X_i & \text{for } i < N \\ C & \text{for } i \geq N. \end{cases}$$

The time homogeneous process X_t which is at a point $x < b$ at time 0, will at a later time Δ have crossed the threshold with some probability. We denote

$$\mathbb{P}_\theta(T_b \leq \Delta | X_0 = x) = G_\theta^b(\Delta|x) \, = \, \int_0^\Delta g_b(r|x)\, dr \tag{2}$$

where $g_b(r|x)$ is the density of T_b given that $X_0 = x$. It may stay below the threshold at all times before Δ, and at time Δ be in state $X_\Delta = y < b$ with probability "density"

$$f_\theta^b(y, \Delta|x) = \frac{\partial}{\partial u} P_\theta(X_\Delta \leq u \ , \ T_b > \Delta | X_0 = x)\Big|_{u=y}. \tag{3}$$

The word probability density is slightly improper for $f_\theta^b(y, \Delta|x)$ since

$$\int_{E_b} f_\theta^b(y, \Delta|x) dy = 1 - G_\theta^b(\Delta|x).$$

The killed process X_i^k allows for the following properly defined transition density

$$f_\theta^k(y, \Delta|x) = f_\theta^b(y, \Delta|x) \cdot \mathbb{1}_{\{x,y \in E_b\}} + G_\theta^b(\Delta|x) \cdot \mathbb{1}_{\{x \in E_b, y = C\}} + \mathbb{1}_{\{x,y=C\}} \tag{4}$$

w.r.t. the measure $\lambda \oplus \delta_C$ defined on $E_b \cup C$, where λ is the Lebesgue measure on $\mathbb{R}$ and δ_C is the Dirac measure in C. Here $\mathbb{1}_A$ denotes the indicator function of the set A.

Observing the original discretized process X_i sequentially up to the stopping time $N - 1$ is equivalent to observe the killed process X_i^k infinitely, since

from the first visit to C no more information is gained. In both cases we can interpret each observed trajectory as a realization of a single random variable. Indeed, an observation $(x_1, \ldots, x_{n-1})$ may be seen as a sequential sample $X(\omega) = (x_1, \ldots, x_{N-1})$ with random length $N(\omega) - 1 = n - 1$ from the process X_i, but also as a single realization of a random variable whose values are $(n-1)$-tuples of elements of the state space E_b for some $n-1$, free to vary from one to any natural number. Accordingly, we have

$$X : \Omega \longrightarrow \bigcup_{j=1}^{\infty} (E_b)^j. \tag{5}$$

Alternatively, the same trajectory can be interpreted as an infinite set of observations $X^k(\omega) = (x_1, \ldots, x_{n-1}, C, \ldots)$ from the killed process X^k where C is always observed from position n and onwards. In this sense it is a single realization of a random variable

$$X^k : \Omega \longrightarrow (E_b \cup C)^{\infty}. \tag{6}$$

Whichever interpretation is preferred, the likelihood of $X(\omega) = (x_1, \ldots, x_{N-1})$ is given by

$$L(X; \theta) = \prod_{i=1}^{\infty} f_\theta^k(x_i, \Delta | x_{i-1}) = \sum_{n=1}^{\infty} \prod_{i=1}^{n-1} f_\theta^b(x_i, \Delta | x_{i-1}) \cdot G_\theta^b(\Delta | x_{n-1}) \cdot \mathbb{1}_{\{N=n\}} \tag{7}$$

for $x_0 \in E_b$ and $(x_1, \ldots, x_n) \in (E_b \cup C)^n$. However, in the sense of (5) it is a density w.r.t. the dominating measure $\oplus_{i=1}^{\infty} \lambda^i$ where λ^i are Lebesgue measures on $\mathbb{R}^i$, while in the sense of (6) it is a density w.r.t. the infinite power $(\lambda \oplus \delta_C)^{\infty}$.

The functions $f_\theta^b(y, \Delta|x)$ and $G_\theta^b(\Delta|x)$ are not generally known in explicit form, except for Brownian motion killed at a constant threshold and a few other cases. Nevertheless, when high frequency data are available, i.e. when Δ is small, some methods that were introduced for different purposes may be applied to get a reliable approximation, which is computationally simple enough to be implemented within an optimization algorithm. That is the subject of the following two Subsections.

2.1 Transition density for small sampling intervals

The transition density and $f_\theta^b(y, \Delta|x)$ are related by the following equation

$$f_\theta^b(y, \Delta|x) = f_\theta(y, \Delta|x) \cdot (1 - \mathbb{P}(T_b < \Delta \mid x, y)) \tag{8}$$

where $\mathbb{P}(T_b < \Delta \mid x, y)$ denotes the probability that the process X conditioned upon being in x at some time t and in y at $t + \Delta$ crosses the threshold for the first time between those two points. In Baldi and Caramellino (2002); Giraudo and Sacerdote (1999) two computationally efficient methods were proposed and compared to approximate this probability when Δ is small.

The main goal was to design a simulation scheme for killed processes (Appendix A.1). Both methods can be applied to our problem. We choose that of Baldi and Caramellino (2002) since it leads to a faster numerical evaluation. This is particularly important since we need to locate the maximum of the likelihood function by means of a numerical optimization algorithm that requires multiple evaluations. Assume the drift and diffusion functions in (1) to be $\mu \in C^1(\mathrm{int}(E_b))$, $\sigma \in C^2(\mathrm{int}(E_b))$ and $\sigma(z) > 0$ for $z \in \mathrm{int}(E_b)$. For $x, y \in E_b$ and small Δ we apply the following approximation

$$\mathbb{P}(T_b < \Delta \mid x, y) \approx \exp\left(-\frac{2}{\Delta}\int_x^b \frac{dz}{\sigma(z)} \cdot \int_y^b \frac{dz}{\sigma(z)}\right) \cdot (1 + \Delta\phi_b)$$

where

$$\phi_b = \begin{cases} \frac{1}{2}\left(\frac{\int_x^y \frac{\lambda(z)dz}{\sigma(z)}}{\int_x^y \frac{dz}{\sigma(z)}} - \frac{\int_x^b \frac{\lambda(z)dz}{\sigma(z)} + \int_y^b \frac{\lambda(z)dz}{\sigma(z)}}{\int_x^b \frac{dz}{\sigma(z)} + \int_y^b \frac{dz}{\sigma(z)}}\right) & \text{if } y \neq x \\ \frac{1}{2}\left(\lambda(y) - \frac{\int_y^b \frac{\lambda(z)dz}{\sigma(z)}}{\int_y^b \frac{dz}{\sigma(z)}}\right) & \text{if } y = x \end{cases}$$

and

$$\lambda(y) = \sigma(y) \cdot \left(\frac{\mu(y)}{\sigma(y)} - \frac{\sigma'(y)}{2}\right)' + \left(\frac{\mu(y)}{\sigma(y)} - \frac{\sigma'(y)}{2}\right)^2$$

where we assume $\lambda(\cdot)$ locally Lipschitz continuous on $\mathrm{int}(E_b)$.

2.2 Conditional distribution of the first-hitting time for small sampling intervals

The density $g_b(r|x)$ satisfies the following integral equation (Buonocore et al., 1987),

$$g_b(r|x) = -2\Psi_b(r|x) + 2\int_0^r g_b(\tau|x)\,\Psi_b(r-\tau|b)\,d\tau, \tag{9}$$

where the function $\Psi_b(r|x)$ is defined as follows,

$$\Psi_b(r|x) = \frac{\partial}{\partial r}\mathbb{P}(X_r < b \mid X_0 = x) + \frac{1}{2}\left(\mu(b) - \frac{1}{4}\sigma'(b)\right) f_\theta(X_r = b \mid X_0 = x).$$

A crude approximation for small r that proves surprisingly efficient in practice is to neglect the integral in (9) and approximate

$$\begin{aligned} G_\theta^b(\Delta|x) &= \int_0^\Delta g_b(r|x)\,dr \approx -2\int_0^\Delta \Psi_b(r|x)\,dr \\ &= 2\,\mathbb{P}(X_r > b \,|X_0 = x) - \left(\mu(b) - \frac{1}{4}\sigma'(b)\right)\int_0^\Delta f_\theta(X_r = b \mid X_0 = x)dr \end{aligned} \tag{10}$$

allowing for reasonably fast evaluations in the optimization algorithm (Appendix A.3). More precise approximations may be achieved (Sacerdote and Tomassetti, 1996), but they require much heavier computational efforts.

3 Examples

Here we present in details a few examples showing that the general theory can effectively be applied for processes which are of practical relevance in many fields. For each example a simulation study is performed to evaluate the quality of the estimators. We generate 10,000 samples of one trajectory, stopped at the first crossing time of a threshold. Details on how to properly detect the first-passage times of discretely simulated paths are given in Appendix A.1. For each trajectory estimates are computed by numerical maximization of the log-likelihood function. A few numerical details are provided in Appendix A.

3.1 Wiener with positive drift

The WD is one of the few cases where (2) and (3) are explicitly known for any value of Δ (the high frequency data assumption is not needed in this case). Consider X_t given by

$$dX_t = \mu dt + \sigma dW_t \tag{11}$$

with μ and σ positive and $X_0 = x_0 < b$. We have (Sacerdote and Giraudo, 2012)

$$f_\theta^b(y, \Delta|x) = f_\theta(y, \Delta|x) \cdot \left[1 - \exp\left(-\frac{2\,(b-y)(b-x)}{\sigma^2 \Delta}\right)\right] \tag{12}$$

and

$$G_\theta^b(\Delta|x) = \frac{1}{2}\left\{1 - \operatorname{erf}\left(\frac{b-x-\mu\Delta}{\sqrt{2\Delta}\,\sigma}\right) + \exp\left(\frac{2\mu(b-x)}{\sigma^2}\right)\left[1 - \operatorname{erf}\left(\frac{b-x+\mu\Delta}{\sqrt{2\Delta}\,\sigma}\right)\right]\right\}$$

where $\operatorname{erf}(\cdot)$ is the error function.

The initial point is fixed at $x_0 = 0$, the threshold is at $b = 10$ and the simulation step is $\Delta = 1$. Results are presented in Table 1. The drift parameter μ is overestimated up to 200% with increasing bias when adding more noise. The distribution of $\hat{\mu}$ is skewed with a long right tail.

3.2 Ornstein-Uhlenbeck process

The OU process is used as a model for many phenomena in physics, biology, engineering and finance. In particular, it is used in neuroscience as the most tractable version of the Leaky Integrate-and-Fire models to represent the evolution of the membrane potential of a neuron between two spikes. A spike is generated whenever the process reaches a *firing threshold* and then the membrane potential is reset immediately to a fixed value x_0 (the *resetting potential*). Then the evolution starts anew with the same law independently of the past. The OU process is solution to the stochastic differential equation

$$dX_t = (-\beta X_t + \mu)\, dt + \sigma dW_t \tag{13}$$

Table 1: WD model. Results of the simulation study. Value used are $x_0 = 0$, $b = 10$ and $\Delta = 1$.

	par	true	avg($\hat{\theta}$)	$\frac{\text{avg}(\hat{\theta})-\theta}{\theta}$	(2.5%, 97.5%)	avg(N)
CASE 1	μ	0.3	0.326	0.086	(0.180, 0.541)	33.72
	σ	0.5	0.488	-0.024	(0.365, 0.612)	
CASE 2	μ	0.3	0.520	0.734	(0.088, 1.631)	34.28
	σ	1.5	1.445	-0.037	(0.932, 1.912)	
CASE 3	μ	0.1	0.125	0.247	(0.045, 0.274)	100.73
	σ	0.5	0.496	-0.009	(0.419, 0.576)	
CASE 4	μ	0.1	0.320	2.203	(0.019, 1.281)	101.61
	σ	1.5	1.467	-0.022	(1.048, 1.844)	

with β and σ positive. It is gaussian with conditional mean and variance

$$\mathbb{E}(X_t|X_0 = x_0) = x_0 \mathrm{e}^{-\beta t} + \frac{\mu}{\beta}\left(1 - \mathrm{e}^{-\beta t}\right), \quad \mathrm{Var}(X_t|X_0 = x_0) = \frac{\sigma^2}{2\beta}\left(1 - \mathrm{e}^{-2\beta t}\right). \tag{14}$$

Thus, $\mathbb{E}(X_\infty) = \mu/\beta$. A threshold is imposed at level $b > x_0 = 0$. The mean first-passage time through the threshold is finite and can be calculated according to formulas in Siegert (1951); Ricciardi and Sato (1988).

Parameter estimation for this model from neuronal data was performed for example in Lansky et al. (2006, 2010); Picchini et al. (2008) by maximizing the product of the gaussian transition densities of the underlying unconstrained OU process. This we will call *naive maximum likelihood*, and it neglects the role of the threshold, and thus, the likelihood is misspecified. In Bibbona et al. (2008) it was shown that considerable bias may occur by estimating in this way.

The likelihood function (7) can be approximated as in Sections 2.1 and 2.2. We have

$$\mathbb{P}(T_b < \Delta \mid x, y) \approx \exp\left(-\frac{2\,(b-x)(b-y)}{\sigma^2 \Delta}\right) \cdot (1 + \Delta\phi_b) \tag{15}$$

with

$$\phi_b = -\frac{\beta\,(b-x)(b-y)(\beta b + \beta x + \beta y - 3\mu)}{3\,\sigma^2\,(2b - x - y)}$$

to be substituted in formula (8), and $G^b_\theta(\Delta|x)$ may be derived evaluating formula (10). The function $\Psi_b(r|x)$ was calculated explicitly in Buonocore et al. (1987), but its numerical integration takes much longer time than the evaluation of the normal cumulative distribution function in the first term and numerical integration in the second summand of the right hand side of (10).

In Table 2 we present the results of a simulation study with 4 different parameter settings, in which we evaluate the performance of the MLE for (β, μ, σ) for the fully specified model and compare it with the naive estimation.

Table 2: OU model. Results of the simulation study. θ indicates either μ, β or σ. Values used are $x_0 = 0$ and $b = 10$.

	true values	method	par	avg($\hat{\theta}$)	$\frac{\text{avg}(\hat{\theta})-\theta}{\theta}$	sd($\hat{\theta}$)	(2.5%, 97.5%)	other values
CASE 1	$\mu = 0.43$ $\beta = 0.05$ $\sigma = 1.2$	Likelihood	μ	0.753	0.752	0.509	(0.207, 2.161)	$\mathbb{E}(X_\infty)$=8.6 $\Delta = 0.1$ $\mathbb{E}(N) = 417.8$
			β	0.081	0.625	0.077	(0.000, 0.276)	
			σ	1.198	-0.001	0.053	(1.091, 1.303)	
		naive	μ	0.779	0.813	0.532	(0.207, 2.254)	
			β	0.091	0.815	0.081	(0.000, 0.294)	
			σ	1.197	-0.003	0.053	(1.087, 1.301)	
CASE 2	$\mu = 1$ $\beta = 0.025$ $\sigma = 1$	Likelihood	μ	1.339	0.339	0.600	(0.584, 2.822)	$\mathbb{E}(X_\infty)$=40 $\Delta = 0.1$ $\mathbb{E}(N) = 114.1$
			β	0.063	1.550	0.082	(0.000, 0.277)	
			σ	0.993	-0.007	0.070	(0.855, 1.135)	
		naive	μ	1.373	0.373	0.630	(0.581, 2.915)	
			β	0.077	2.062	0.089	(0.000, 0.300)	
			σ	0.990	-0.010	0.071	(0.851, 1.133)	
CASE 3	$\mu = 2$ $\beta = 0.2$ $\sigma = 1.7$	Likelihood	μ	2.638	0.319	1.370	(0.931, 6.150)	$\mathbb{E}(X_\infty)$=8.33 $\Delta = 0.1$ $\mathbb{E}(N) = 126.8$
			β	0.282	0.173	0.218	(0.000, 0.790)	
			σ	1.686	-0.008	0.126	(1.428, 1.933)	
		naive	μ	2.783	0.392	1.445	(0.967, 6.457)	
			β	0.322	0.344	0.229	(0.000, 0.869)	
			σ	1.679	-0.012	0.127	(1.420, 1.929)	
CASE 4	$\mu = 8$ $\beta = 1$ $\sigma = 1$	Likelihood	μ	8.151	0.019	1.315	(5.775, 10.934)	$\mathbb{E}(X_\infty)$=8 $\Delta = 0.49$ $\mathbb{E}(N) = 121.6$
			β	1.003	0.003	0.177	(0.657, 1.361)	
			σ	1.011	0.011	0.108	(0.814, 1.253)	
		naive	μ	8.326	0.041	1.322	(5.944, 11.156)	
			β	1.033	0.033	0.175	(0.698, 1.393)	
			σ	0.984	-0.016	0.114	(0.734 1.205)	

When the asymptotic level $\mathbb{E}(X_\infty) = \mu/\beta$ is above b is denoted supra-threshold regime, while sub-threshold regime when it is below. In Case 1 parameters are chosen in sub-threshold regime with values compatible with those expected for the membrane potential of a neuron during spontaneous activity (Lansky et al. (2006), values are expressed in units of milliseconds and millivolts). In Case 2 they are chosen according to the ones estimated in Lansky et al. (2010) during stimulation, yielding supra-threshold dynamics. Cases 3 and 4 are illustrative of different ranges. In case 4 a larger simulation step is used to test if the approximation is still acceptable.

The two methods provide similar estimates, but the naive method performs slightly worse. The superiority of the MLE will become apparent in Section 5.1. Table 2 shows that $\hat{\mu}$ displays a skewed distribution with a heavy right tail, and that $\hat{\beta}$ is relatively more variable compared to the other two parameters. Estimates of σ are good whatever method is applied. In Bibbona et al. (2008) the bias incurred in $\hat{\mu}$ from naive estimation of μ and σ when β is assumed known, was evaluated approximately to be σ^2/b. This value largely underestimates the true magnitude of the bias when β also has to be estimated.

3.3 Square root process

The SR process is the solution to the following stochastic differential equation

$$dX_t = (-\beta X_t + \mu)\, dt + \sigma\sqrt{X_t} dW_t \tag{16}$$

with β and σ positive. We assume $2\mu \geq \sigma^2$, in which case the boundary 0 is entrance, following Feller's classification. The SR process was first studied in pure mathematics (Feller, 1951) as an example of a singular diffusion process, and it gained popularity under the name of the CIR process in finance as a model for interest rates (Cox et al., 1985). In neuroscience it has been proposed as a Leaky Integrate-and-Fire model for the membrane potential evolution between two spikes (Giorno et al., 1988). It is considered more realistic than the OU process since it is bounded from below. The transition density is a non-central chi-square distribution with non centrality parameter

$$\lambda = \frac{4x_0\beta \mathrm{e}^{-\beta t}}{\sigma^2(1-\mathrm{e}^{-\beta t})}$$

and $k = 4\mu/\sigma^2$ degrees of freedom. The mean first-passage time through a constant boundary b is finite; an explicit expression is given in Lansky et al. (1995).

The likelihood function (7) can be approximated as in Sections 2.1 and 2.2. In particular,

$$\mathbb{P}(T_b < \Delta \mid x, y) \approx \exp\left(-\frac{8\,(\sqrt{b}-\sqrt{x})(\sqrt{b}-\sqrt{y})}{\sigma^2\Delta}\right)\cdot(1+\Delta\phi_b) \tag{17}$$

Table 3: SR model. Results of the simulation study. θ indicates either μ, β or σ.

	par	true	avg($\hat{\theta}$)	$\frac{\text{avg}(\hat{\theta})-\theta}{\theta}$	sd($\hat{\theta}$)	(2.5%, 97.5%)	x_0	b	$\mathbb{E}(X_\infty)$	Δ	$\mathbb{E}(N)$
	μ	10	11.874	0.187	7.363	(2.240, 29.494)					
CASE 1	β	1.2	1.331	0.109	0.999	(0.000, 3.571)	5	10	8.3	0.08	51.4
	σ	0.7	0.686	-0.019	0.093	(0.493, 0.875)					
	μ	6	7.154	0.192	4.416	(1.607, 17.945)					
CASE 2	β	0.31	0.351	0.131	0.290	(0.000, 1.021)	10	20	19.4	0.12	62.8
	σ	0.5	0.490	-0.018	0.052	(0.386, 0.595)					
	μ	2	4.054	1.027	3.356	(0.863, 13.198)					
CASE 3	β	0.05	0.163	2.257	0.214	(0.000, 0.730)	10	20	40	0.08	96.4
	σ	0.5	0.495	-0.010	0.041	(0.412, 0.576)					

with

$$\begin{aligned}\phi_b = -\Big\{ &\Big[\sqrt{x}\,(b-y)\left(y\beta^2 b - 3\sigma^2\mu + \frac{9}{16}\sigma^4 + 3\mu^2\right)\Big] \\ &-\Big[\sqrt{y}\,(b-x)\left(x\beta^2 b - 3\sigma^2\mu + \frac{9}{16}\sigma^4 + 3\mu^2\right)\Big] \\ &-\Big[\sqrt{b}\,(x-y)\left(xy\beta^2 - 3\sigma^2\mu + \frac{9}{16}\sigma^4 + 3\mu^2\right)\Big]\Big\} \\ &\Big/\Big[3\,\sigma^2\sqrt{xyb}\,(\sqrt{x}-\sqrt{y})\,(-2\sqrt{b}-+\sqrt{x}+\sqrt{y})\Big]\end{aligned}$$

to be substituted in formula (8), and $G^b_\theta(\Delta|x)$ may be derived evaluating formula (10). The function $\Psi_b(r|x)$ was calculated explicitly in Giorno et al. (1989). As in the OU case, its numerical integration takes much longer time than the evaluation of the right hand side of (10) using the cumulative non-central chi-square distribution function in the first term and numerical integration in the second.

In Table 3 we present the results of a simulation study in three different parameter settings. A short account on a few numerical difficulties is given in Appendix A.3. The choice of parameters is representative of the two main regimes: sub-threshold (Cases 1 and 2) and supra-threshold (Case 3). Results are similar to those in the previous examples.

3.4 Bootstrap bias correction

A common feature emerging from the examples is that MLE for killed processes may be heavily biased, largely variable and skewed. In some cases the bias is an effect of the sequential sampling: MLE is unbiased for the same model observed with a fixed length, even if short. It is the case of the WD, but also of the OU process when the parameter β is known and only μ and σ are estimated (Bibbona et al., 2008) and of some discrete models such as multidimensional

random walks (Bibbona and Rubba, 2011). The same "optional sampling effect" was noticed in MLE following a sequential test (Whitehead, 1986), in clinical trials (Jung and Kim, 2004) and is known for binomial samples since Girshick et al. (1946). For the OU and SR processes, on the contrary, even with a standard sampling scheme where trajectories are recorded up to a fixed but small time, bias, large variance and skewness would be equally present. This phenomenon was illustrated for example in Tang and Chen (2009) where parametric bootstrap was shown to be effective in reducing the bias of the estimates for diffusion processes observed up to a short fixed time without increasing the variance too much. We will test the same method for our sequential problem. First we generate our simulated sample with parameter θ_0 (3,000 trajectories). For each path we first calculate the MLE $\hat{\theta}$ and then we generate a bootstrap sample simulating 1,000 trajectories with the estimated parameter. On the bootstrap sample we estimate again path by path and calculate the average estimated bootstrap parameter $\text{avg}(\hat{\theta}_B)$. The bias at $\hat{\theta}$ is evaluated as $\text{bias}(\hat{\theta}) = \text{avg}(\hat{\theta}_B) - \hat{\theta}$ and the estimate $\hat{\theta}$ is corrected to $\hat{\theta}^{BC} = \hat{\theta} - \text{bias}(\hat{\theta}) = 2\hat{\theta} - \text{avg}(\hat{\theta}_B)$ assuming that the bias at θ_0 and that at $\hat{\theta}$ are not much different. Results are illustrated in Table 4. The relative efficiency is defined as the ratio between the determinants of the sample mean square error matrices of the two estimators. It is geometrically interpreted as the squared ratio between the area of the concentration ellipsoids of the random vectors $\hat{\theta}^{BC} - \theta_0$ and $\hat{\theta} - \theta_0$ (Cramér (1946)). When it is smaller then one, the bootstrap corrected estimators are more concentrated than their original counterparts. This method provides a better comparison between two families of joint estimators w.r.t. the individual mean squared errors since it takes into account not only the marginal distribution of the estimates, but also their joint behavior. Our simulations confirm that parametric bootstrap is very effective in removing the bias but the increase in variability is not always negligible.

4 Consistency and asymptotic normality

For processes killed at a threshold standard asymptotics do not apply. When the process is killed no further information can be obtained, and the asymptotic scenario of number of observations going to infinity (for fixed sampling interval) is no longer relevant. Moreover, on a fixed interval $[0, n\Delta]$, increasing n by decreasing Δ does not improve estimators of drift parameters, which are inconsistent even for the fully continuously observed trajectory. Asymptotic properties of MLEs may nevertheless be exploited by considering a collection of independent realizations of the killed process and drawing inference from the global log-likelihood function of the entire sample. This kind of asymptotics has already been introduced for regular Markov chains with a finite number of states in Bhat (1961).

Assume a sample of m independent discretely observed trajectories of X. Since every trajectory may be considered as a single independent random variable according to (5) (the choice in the following) or (6), the global log-likelihood

Table 4: Bootstrap. Results of the simulation study. θ indicates either μ, β or σ.

	par	true	avg($\hat{\theta}$)	avg($\hat{\theta}^{BC}$)	sd($\hat{\theta}$)	sd($\hat{\theta}^{BC}$)	rel.eff.
	μ	0.43	0.745	0.431	0.524	0.549	
OU CASE 1	β	0.05	0.080	0.051	0.078	0.095	0.927
	σ	1.2	1.197	1.201	0.052	0.052	
	μ	1	1.324	1.016	0.596	0.644	
OU CASE 2	β	0.025	0.062	0.030	0.081	0.100	1.202
	σ	1	0.992	1.001	0.071	0.072	
	μ	10	12.014	9.887	7.774	8.584	
SR CASE 1	β	1.2	1.352	1.174	1.044	1.181	0.990
	σ	0.7	0.690	0.707	0.093	0.093	
	μ	2	4.112	2.341	3.372	3.946	
SR CASE 3	β	0.05	0.167	0.075	0.215	0.264	1.025
	σ	0.5	0.495	0.500	0.041	0.041	

function can be written as

$$\ell_m(X^{(1)}, \cdots, X^{(m)}; \theta) = \sum_{i=1}^{m} \log L(X^{(i)}; \theta)$$

where $L(X^{(i)}; \theta)$ is the likelihood (7) of a single trajectory. A function $h : E \times \Theta \longrightarrow \mathbb{R}$ is called *locally dominated integrable* w.r.t. a measure μ if for each $\theta' \in \Theta$ there exists a neighborhood $U_{\theta'}$ of θ' and a non-negative μ-integrable function $k_{\theta'} : E \longrightarrow \mathbb{R}$ such that $|h(x, \theta)| \leq k_{\theta'}(x)$ for all $x \in E$ and all $\theta \in U_{\theta'}$.

Standard theorems for i.i.d. samples (Cramér (1946) or the more modern Sørensen (1999); Sørensen (2011)) guarantee that a consistent root of the likelihood equation exists which is asymptotically normal and efficient provided the following condition is fulfilled.

Condition 4.1

1. *The function $L(X; \theta)$ is twice continuously differentiable w.r.t. θ for all $X \in \bigcup_{j=1}^{\infty}(E_b)^j$.*

2. *For every fixed $X \in \bigcup_{j=1}^{\infty}(E_b)^j$, the functions $\partial_{\theta_i} L(X; \theta)$, $\partial_{\theta_i \theta_j} L(X; \theta)$, $i, j \in \{1, \ldots, p\}$, are locally dominated integrable w.r.t. the measure $\oplus_{i=1}^{\infty} \lambda^i$. Moreover, the functions $\partial_{\theta_i \theta_j} \log L(X; \theta)$ are locally dominated integrable w.r.t. the measure induced by $L(X; \theta_0)$.*

3. *The information matrix with elements $m_{ij} = \mathbb{E}_{\theta_0}\left[\partial_{\theta_i} \log L(X; \theta)\, \partial_{\theta_j} \log L(X; \theta)\right]_{\theta=\theta_0}$ is finite and positive definite, where $\mathbb{E}_{\theta_0}(\cdot)$ denotes expectation w.r.t. $\mathbb{P}_{\theta_0}$.*

Conditions 4.1 (1) and 4.1 (2) are easily transferred to the transition density (4), while Condition 4.1 (3) is unmanageable due to the complicated expression

of the likelihood (7). We will show that it is equivalent to a simpler one that only involves the transitions (4).

Under suitable regularity conditions (interchange of integration and differentiation), for all $i \in \{1, \cdots, p\}$ and for any $x \in E_b$, we have

$$\begin{aligned}
&\mathbb{E}_\theta\big[\partial_{\theta_i} \log f_\theta^k(X_{t+\Delta}, \Delta | X_t = x)\big] \\
&\quad = \int_{E_b} \partial_{\theta_i} \log f_\theta^b(y, \Delta|x) f_\theta^b(y, \Delta|x) dy + \partial_{\theta_i} \log G_\theta^b(\Delta|x) G_\theta^b(\Delta|x) \;=\; 0, \quad (18)
\end{aligned}$$

and $\mathbb{E}_\theta\big[\partial_\theta \log L(X;\theta)\big] = 0$ (the null vector) as always happens with regular score functions.

The elements m_{ij} of the information matrix can be calculated as follows:

$$\begin{aligned}
m_{ij} =& \mathbb{E}_\theta\big[\partial_{\theta_i} \log L(X;\theta)\, \partial_{\theta_j} \log L(X;\theta)\big] \\
=& \sum_{n\in\mathbb{N}} \int_{(E_b)^{n-1}} \left[\sum_{a=1}^{n-1} \partial_{\theta_i} \log f_\theta^k(x_a, \Delta|x_{a-1})\right] \left[\sum_{h=1}^{n-1} \partial_{\theta_j} \log f_\theta^k(x_h, \Delta|x_{h-1})\right] d\mathbb{P}_{\theta,n}(X) \\
=& \sum_{n\in\mathbb{N}} \int_{(E_b)^{n-1}} \Bigg[\sum_{a=1}^{n-1} \partial_{\theta_i} \log f_\theta^b(x_a, \Delta|x_{a-1}) \partial_{\theta_j} \log f_\theta^b(x_a, \Delta|x_{a-1}) \\
&\qquad + \partial_{\theta_i} \log G_\theta^b(\Delta|x_{n-1})\, \partial_{\theta_j} \log G_\theta^b(\Delta|x_{n-1})\Bigg] d\mathbb{P}_{\theta,n}(X) \\
=& \sum_{a=1}^{\infty} \int_{(E_b)^2} \partial_{\theta_i} \log f_\theta^b(x_a, \Delta|x_{a-1})\, \partial_{\theta_j} \log f_\theta^b(x_a, \Delta|x_{a-1}) \\
&\qquad \times f_\theta^b\big(x_{a-1}, (a-1)\Delta|x_0\big) f_\theta^b(x_a, \Delta|x_{a-1})\, dx_a\, dx_{a-1} \\
&+ \sum_{n=2}^{\infty} \int_{E_b} \partial_{\theta_i} \log G_\theta^b(\Delta|x_{n-1})\, \partial_{\theta_j} \log G_\theta^b(\Delta|x_{n-1}) f_\theta^b\big(x_{n-1}, (n-1)\Delta|x_0\big) \\
&\times G_\theta^b(\Delta|x_{n-1})\, dx_{n-1} + \partial_{\theta_i} \log G_\theta^b(\Delta|x_0)\, \partial_{\theta_j} \log G_\theta^b(\Delta|x_0)\, G_\theta^b(\Delta|x_0).
\end{aligned}$$

where $d\mathbb{P}_{\theta,n}(X) = f_\theta^b(x_1, \Delta|x_0) \cdots f_\theta^b(x_{n-1}, \Delta|x_{n-2})\, G_\theta^b(\Delta|x_{n-1})\, dx_1 \cdots dx_{n-1}$ and θ is evaluated at θ_0. The second equality holds by definition. The third equality follows from (18) when a and h are different. The fourth equality is obtained inverting the order of the summations and performing integration over all variables except for x_a and x_{a-1} using the following Chapman-Kolmogorov relations

$$\begin{aligned}
\int_{E_b} f_\theta^b(y, t|x) f_\theta^b(x, s|z) dx &= f_\theta^b(y, t+s|z) \\
\int_{E_b} G_\theta^b(t|x) f_\theta^b(x, s|z) dx &= G_\theta^b(t+s|z) - G_\theta^b(s|z),
\end{aligned}$$

and that since the crossing of the threshold is a sure event, we have

$$\sum_{n=a+1}^{\infty} \big[G_\theta^b\big((n-a)\Delta|x_a\big) - G_\theta^b\big((n-a-1)\Delta|x_a\big)\big] = 1.$$

We rename the variables x_{a-1} and x_a in the right hand side of the last equality of m_{ij} by x and y, using that the process is time homogenous. We therefore have

$$\begin{aligned} m_{ij} =& \int_{(E_b)^2} \partial_{\theta_i} \log f^b_\theta(y,\Delta|x) \partial_{\theta_j} \log f^b_\theta(y,\Delta|x) f^b_\theta(y,\Delta|x) \sum_{n=0}^{\infty} f^b_\theta\big(x, n\Delta|x_0\big)\, dy\, dx \\ &+ \int_{E_b} \partial_{\theta_i} \log G^b_\theta(\Delta|x)\, \partial_{\theta_j} \log G^b_\theta(\Delta|x) G^b_\theta(\Delta|x) \sum_{n=1}^{\infty} f^b_\theta\big(x, n\Delta|x_0\big) dx \\ &+ \partial_{\theta_i} \log G^b_\theta(\Delta|x_0)\, \partial_{\theta_j} \log G^b_\theta(\Delta|x_0) G^b_\theta(\Delta|x_0). \end{aligned} \tag{19}$$

For the matrix elements m_{ij} to be finite we need $\sum_{n=0}^{\infty} f^b_{\theta_0}\big(x, n\Delta|x_0\big)$ to converge to a function $\nu(x)$. Then $\int_{E_b} \nu(x)dx = \sum_{n=0}^{\infty} \int_{E_b} f^b_{\theta_0}(x, n\Delta|x_0)dx = \sum_{n=0}^{\infty} P(N > n\Delta)$. Thus, $\mathbb{E}_{\theta_0}(N) = \sum_{n=0}^{\infty} P(N \geq n\Delta) = \int_{E_b} \nu(x)dx + \sum_{n=1}^{\infty} P(N = n)$. Since the passage is a sure event, the normalized function

$$\pi_\theta(x) = \frac{\sum_{n=0}^{\infty} f^b_\theta\big(x, n\Delta|x_0\big) \cdot \mathbb{1}_{\{x \in E_b\}} + \mathbb{1}_{\{x=C\}}}{\mathbb{E}_\theta(N)}$$

is a probability density on $E_b \cup C$, and the function

$$Q^\Delta_\theta(x,y) = \pi_\theta(x) f^k_\theta(y,\Delta|x)$$

is a joint probability density on $(E_b \cup C)^2$. Finiteness of $\mathbb{E}_{\theta_0}(N)$ and convergence of the above series are equivalent. Thus, if $\mathbb{E}_{\theta_0}(N) < \infty$ (and so $\mathbb{E}_{\theta_0}(T_b) < \infty$) we can recast formula (19) into the following form

$$m_{ij} = \int_{(E_b \cup C)^2} \partial_{\theta_i} \log f^k_\theta(y,\Delta|x)\, \partial_{\theta_j} \log f^k_\theta(y,\Delta|x)\, Q^\Delta_\theta(x,y)\, dy\, dx. \tag{20}$$

Finiteness of m_{ij} is now restated as the square integrability of $\partial_\theta \log f^k_\theta(y,\Delta|x)$ evaluated at θ_0 w.r.t. the measure with density $Q^\Delta_{\theta_0}(x,y)$.

We have proved that the following condition is equivalent to Condition 4.1.

Condition 4.2

1. *$f^k_\theta(x,y)$ is twice continuously differentiable w.r.t. θ for all $(x,y) \in (E_b \cup C)^2$.*
2. *For every fixed $y \in E_b$, the functions $\partial_{\theta_i} f^k_\theta(x,y)$, $\partial_{\theta_i\theta_j} f^k_\theta(x,y)$, $i,j \in \{1,\dots,p\}$, are locally dominated integrable w.r.t. the measure $\lambda + \delta_C$. Moreover, the functions $\partial_{\theta_i\theta_j} \log f^k_\theta(x,y)$ are locally dominated integrable w.r.t. $Q^\Delta_{\theta_0}(x,y)$.*
3. *$\mathbb{E}_{\theta_0}(T_b) < \infty$.*
4. *The information matrix with elements (20) is finite and positive definite.*

The probabilistic meaning of (20) and Condition 4.2 is as follows.

Remark 4.1 *The probability density $\pi_{\theta_0}(x)$ can be interpreted as the density of the stationary measure of a regenerative process which coincides with X_i^k up to the first time N it is in C, at the next step it is reset to x_0 with probability one, and then it starts anew with the same law (Meyn and Tweedie, 2009, Thm. 10.0.1). Indeed, Condition 4.2* (3) *ensures that such a process is stationary and that $\pi_{\theta_0}(x)$ is well defined. $Q_{\theta_0}^{\Delta}(x, y)$ coincides with the joint stationary density of two consecutive observations from the regenerative process whenever $x \in E_b$. Asymptotic properties for many trajectories could have been derived directly for a sample of such a stationary regenerative process which is observed for a long time under the same hypotheses (Sørensen, 1999; Sørensen, 2011).*

For diffusion processes the transition densities are often not known in explicit form. Parameter estimation could then be based on some martingale estimating functions which play the role of the score vector. If we can find a martingale estimating function for the killed process X_i^k, a straightforward modification of Condition 4.2 still guarantees that a consistent root of the estimating equation exists, which is asymptotically normal when the number of trajectories becomes large.

5 Examples with multiple trajectories

Based on the asymptotic results in the previous Section, one might be encouraged to use MLE to estimate from a sample of multiple trajectories. However, the number of trajectories is always finite, and when sampling from a diffusion process, the sampling interval cannot be as small as we like. To assess the validity of the approximations introduced for non-trivial diffusions, we return to simulations. The simulated samples from Section 2 are used again, but now the samples are divided in groups of m trajectories. We estimate from each group by maximizing its global likelihood. Different values of m are used (in particular $m = 1, 3, 10, 30, 100$) to show which is the number of trajectories needed to get reliable results. We only present the most relevant cases. In particular, WD is not illustrated since all effects are visible in the more interesting models.

5.1 OU model

In the upper panels of Figure 1, density plots of the relative estimates $(\hat{\theta}_m - \theta)/\theta$ are displayed, where $\hat{\theta}_m$ is the estimated value from a group of m trajectories. The true parameters are those of Case 1 in Table 2. When estimating from single trajectories, the probability of getting an estimate which differs from the true value by more than 100% is high. Increasing m the distribution of the estimates becomes more symmetric and concentrated around the true value. Even for $m = 3$ or 10, the quality of the estimator is considerably increased, and when $m = 30$ or 100, parameters are well estimated.

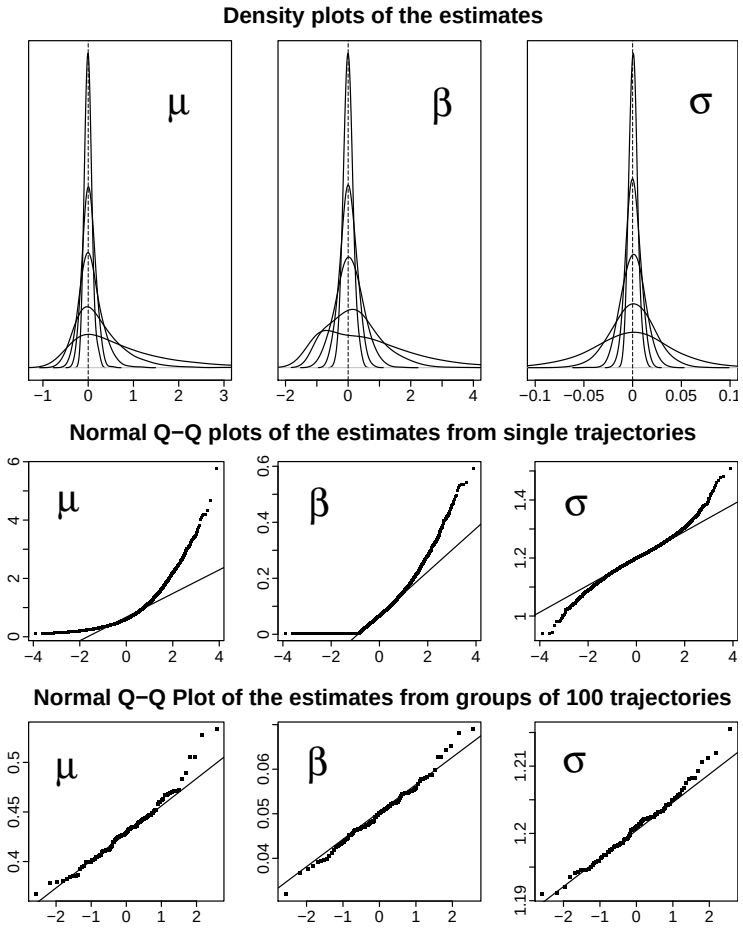

Figure 1: OU model, Case 1. Upper panels: density plots of the relative estimates $(\hat{\theta}_m - \theta)/\theta$, where $\hat{\theta}_m$ is the estimated value from a group of m trajectories, for $m = 1, 3, 10, 30$ and 100 (from less to more peaked). Lower panels: normal Q-Q plots of the estimates, from single trajectories (middle) and from groups of 100 trajectories (lower panels).

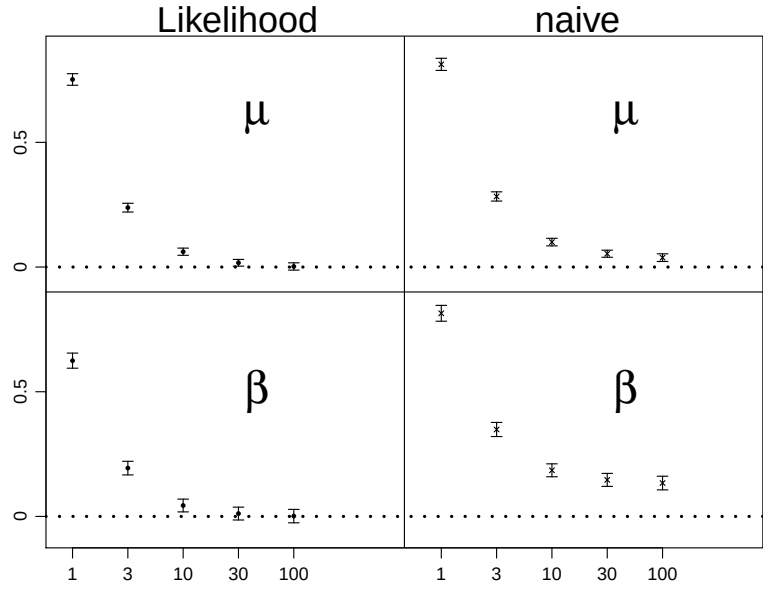

Figure 2: OU model, Case 1. Confidence intervals for the mean relative bias $(\hat{\theta} - \theta)/\theta$ of MLE and naive estimates by groups of trajectories, against m, the number of trajectories used in each group.

In the lower panels of Figure 1, normal Q-Q plots of the estimates are displayed, first from single trajectories and then from groups of 100 trajectories. While we are far from normality in the first case, an approximately normal distribution is achieved for $m = 100$.

Figure 2 illustrates confidence intervals for the mean relative bias for the drift parameters of the MLE and of the naive method. They are calculated according to the following formula

$$\frac{\text{avg}(\hat{\theta}_m) - \theta}{\theta} \pm \frac{t_{(0.975,k-1)}\text{sd}(\hat{\theta}_m)}{\theta\sqrt{k}}$$

where m is the number of trajectories per group, and the integer $k \sim 10,000/m$ is the number of repetitions. Here, $t_{(0.975,k-1)}$ is the 97.5% quantile of the t-distribution with $k-1$ degrees of freedom. This approximation formula holds when $m = 100$ since we have approximately normal estimates and $k = 100$ groups, while when $m = 1$ the estimates are far from normal, but the sample is now large ($k = 10,000$).

If the true likelihood is used, the large bias in the estimated drift parameters from single trajectories already disappears for $m = 30$. The amplitude of the confidence interval does not change much when we use large numbers of trajectories, since the reduction of the variance of $\hat{\theta}_m$ is compensated by the increase in the factor $1/\sqrt{k}$.

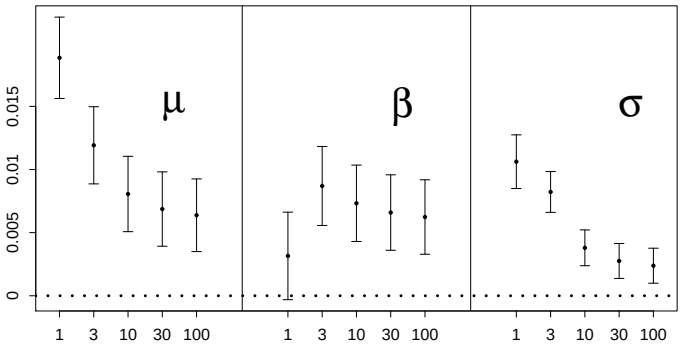

Figure 3: OU model. Case 4. Confidence intervals of the mean relative bias $(\hat{\theta}-\theta)/\theta$ estimated from groups of trajectories with a non-infinitesimal sampling step ($\Delta = 0.49$) against m, the number of trajectories used in the estimation.

Increasing m it becomes apparent that the naive estimators for the drift parameters settle to some constant level, which does not always coincide with the true values: $\hat{\beta}$ differs from β by 12% in the case plotted.

The quality of the estimator is sensitive to the approximations introduced. Consider Case 4 in Table 2, where the step of the discretization is $\Delta = 0.49$, and thus not small. Density plots and normal Q-Q plots are similar to those illustrated for Case 1 (not shown). Confidence intervals of the mean relative bias are plotted in Figure 3. Parameters are chosen such that the mean of the sample size N in each trajectory is comparable to the other cases, and thus, the observation intervals ($N\Delta$) are longer. The estimates of the drift parameters from single trajectories are consequently better. When estimating from many trajectories, however, some very small asymptotic bias is now visible even when estimating from our best approximation of the likelihood. Nevertheless, the method is robust and the bias is less than 1%.

To conclude, 30 trajectories are enough to get approximately unbiased estimates, and for $m = 100$, the estimator is also approximately normally distributed.

5.2 Square root model

Despite minor numerical complications (Appendix A.3), the SR model behaves essentially like the OU model, cf. Figure 4. The main difference is that the estimator of σ has larger variance. Again 30 trajectories are enough for unbiasedness and normality is almost achieved with $m = 100$.

6 Intracellular recordings from a motoneuron

The membrane potential from a spinal motoneuron in segment D10 of an adult red-eared turtle (*Trachemys scripta elegans*) was recorded while a periodic me-

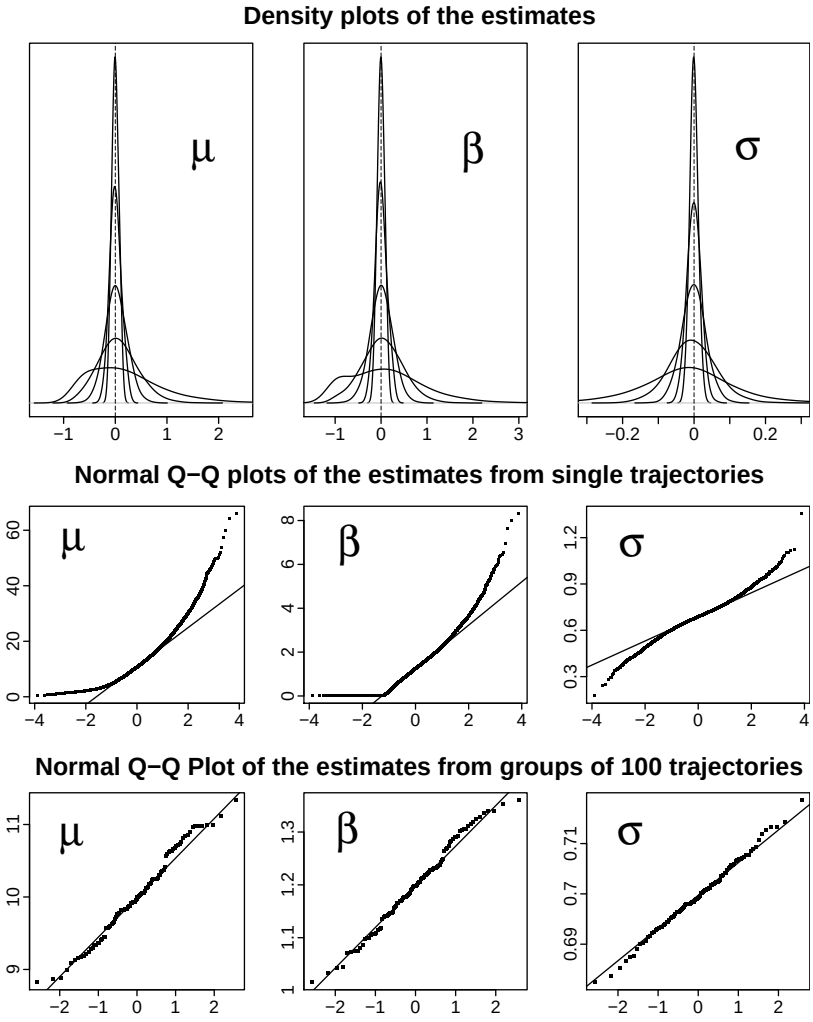

Figure 4: SR model, Case 1. Panels as in Fig. 1.

chanical stimulus was applied to selected regions of the carapace with a sampling step of 0.1 ms (for details see Berg et al. (2007, 2008)). The turtle responds to the stimulus with a reflex movement of a limb known as the *scratch reflex*, causing an intense synaptic input to the recorded neuron. Due to the time varying stimulus, a model for the complete data set needs to incorporate the time-inhomogeneity, as done in Jahn et al. (2011). The data can only be assumed stationary during short time windows, which is required for the models introduced in Sections 3.2 and 3.3, called Leaky Integrate-an-Fire (LIF) models in computational neuroscience. Here, the crossing of the threshold corresponds to a firing of the neuron (a spike), and a trajectory corresponds to an interspike interval. In the LIF models the intensity of the synaptic input is given by the parameters μ and σ, and under stimulation they are both higher, causing the mean membrane potential to increase. Also β, the inverse of the time constant, depends on the conductance and is state dependent. Thus, the parameters are influenced by the intensity of the input and cannot be considered constant along the experiment since the stimulation is varying. During intense stimulation, however, the neuron fires frequently and the typical length of an interspike interval can be assumed smaller than the time scale of the variability of the input. Therefore, we consider the input approximately constant for at least 3-4 consecutive trajectories during on-cycles (following Jahn et al. (2011)). From non-parametric estimation it was shown that after a translation of the data which set the zero level of the potential (the inhibitory reversal potential) to approximatively -74.5 mV, a LIF model based on a SR process describes the data in the on-cycles better than the OU model (Jahn et al. (2011)). Spikes are easily identified and the beginning of each trajectory was fixed to the first recorded point after the spike that is above -60 mV (15.4 mV after translation). The threshold is accommodated manually just above the highest local maximum of each group of homogeneous consecutive trajectories within the same on-cycle. Parameter estimation is performed path by path both with the full MLE and with the naive approximation. The analyzed data are plotted in Fig. 5 and results are presented in Table 5 together with a global estimate from all 3-4 consecutive trajectories as a unique sample according to the method in Section 4. The sample is small and the asymptotic regime is still not reached, but according to the simulation results in Section 5.2 the quality of the estimate is substantially improved. For these data, the naive estimates do not differ much from the MLEs, probably because the trajectories are relatively long, on average 342 points corresponding to 34.2 ms. The individual estimates done trajectory by trajectory show large statistical fluctuations, which should be reduced in the global estimates, as well as the small sample bias.

7 Conclusion

Drawing inference for processes which are killed the first time they cross a threshold revealed an intriguing task. In the neurobiological literature where the problem is often encountered, estimation has always been performed ig-

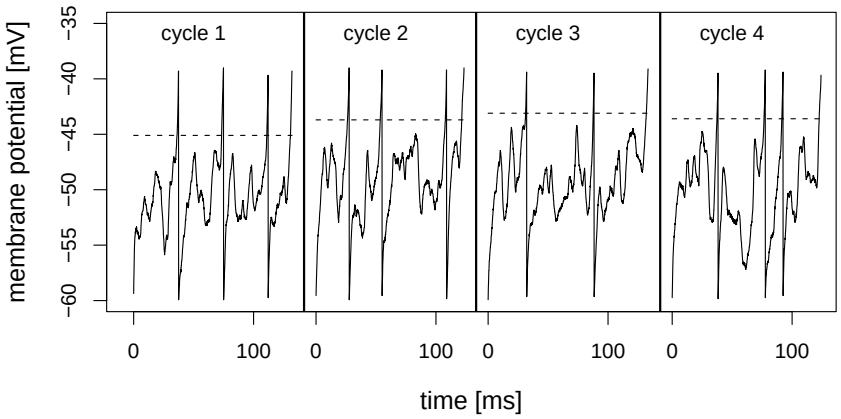

Figure 5: Four cycles of intracellular recordings of the membrane potential of a motoneuron during mechanical stimulation obtained from an isolated carapace-spinal cord preparation from adult turtles. The dashed lines are the thresholds used in the analysis.

Table 5: Experimental data from a motoneuron fitted with a SR model. The starting point x_0 is different trajectory by trajectory, but always around 15.5 mV. The naive estimate of σ always coincides with the MLE.

		$\hat{\mu}$		$\hat{\beta}$		$\hat{\sigma}$	$\hat{\mu}_m$	$\hat{\beta}_m$	$\hat{\sigma}_m$	b
cycle	traj	MLE	naive	MLE	naive			global MLE		
1	1	3.53	3.57	0.132	0.134	0.118	4.43	0.164	0.108	29.4
	2	3.14	3.19	0.112	0.114	0.096				
	3	5.83	5.96	0.221	0.227	0.104				
	4	8.35	8.57	0.317	0.327	0.112				
2	1	6.88	6.91	0.242	0.243	0.110	5.76	0.203	0.108	30.8
	2	4.35	4.56	0.155	0.164	0.117				
	3	4.92	4.97	0.177	0.178	0.100				
	4	9.73	9.80	0.319	0.322	0.090				
3	1	4.87	4.90	0.167	0.168	0.113	4.18	0.148	0.105	31.4
	2	4.03	4.08	0.149	0.151	0.103				
	3	4.34	4.08	0.151	0.153	0.099				
4	1	6.18	6.28	0.222	0.226	0.118	3.14	0.106	0.119	30.9
	2	1.43	1.50	0.044	0.047	0.124				
	3	1.15	1.12	0.001	0.000	0.100				
	4	6.30	6.44	0.226	0.232	0.114				

noring the presence of the threshold, assuming wrongly that the sample is drawn from an unconstrained process. This approach was already criticized in Bibbona et al. (2008, 2010); Giraudo et al. (2011), but no general method is available in the statistical literature for this problem. We compute the correct likelihood accounting for the presence of the threshold, and show that even if the model is correctly specified, estimates from a single trajectory are poor. Bootstrap bias correction may be applied to improve estimators. Standard asymptotic theory does not apply to a single trajectory. The problem is solved and good estimates can be achieved if we can infer from a large sample made of many trajectories. Asymptotic results are proved in this case, and we show numerical evidence that 30–100 trajectories are enough to get reliable estimates. Even with 3–10 trajectories inference is greatly improved. In neuroscience, sample sizes of 50–1,000 repetitions of inter-spike intervals are common. This emphasizes the importance of knowing how to implement the correct likelihood.

Many interesting questions remain open. In neuroscience applications, even if many trajectories are recorded, it might be questionable to assume that parameters are stable along repetitions. For the data set analyzed in Section 6 the problem is apparent. It would be appropriate to incorporate random effects (Picchini et al., 2008) or to consider more complicated inhomogeneous models (Jahn et al., 2011). Moreover, a more flexible spike generation mechanism would probably be more realistic and different models have been designed to this aim, cf. (Jahn et al., 2011) and references therein. The lesson learned here is that a correct specification of the likelihood function which incorporates the presence of the threshold improve the estimates.

In a few special cases, some ad-hoc methods that were designed for sequential analysis were shown to provide unbiased estimators for a single trajectory of certain killed processes (Girshick et al., 1946; Ferebee, 1983; Bibbona and Rubba, 2011). It would be useful to find unbiased estimators for non-trivial diffusions also, and to provide a detailed comparison with the likelihood approach, especially for those experiments where only one trajectory is available. Moreover, for most continuous time models, the transition density is not available, and the estimating function approach can be a good solution. To this aim suitable estimating functions have to be designed, which might also allow to remove the high frequency assumption. The challenge is to find explicit expressions for conditional moments of suitable functionals from the constrained distribution. Finally, the numerical approximations to the crossing probabilities in the OU and the SR models are only valid for small sampling steps. This is not a problem for neurophysiological data, which are typically high frequency, but in other applications it might be a major drawback.

Acknowledgements

EB dedicates this paper to his newborn daughter Caterina. The authors thank Rune Berg for making his experimental data available. We would also like to thank Laura Sacerdote, Michael Sørensen, Martin Jacobsen and Stefano Iacus

for enlightening discussions and Michela Costanzo for the prompt and helpful technical support on parallel computing.

References

Baldi, P. and L. Caramellino (2002). Asymptotics of hitting probabilities for general one-dimensional pinned diffusions. *Ann. Appl. Probab. 12*(3), 1071–1095.

Berg, R. W., A. Alaburda, and J. Hounsgaard (2007). Balanced inhibition and excitation drive spike activity in spinal halfcenters. *Science 315*, 390–393.

Berg, R. W., S. Ditlevsen, and J. Hounsgaard (2008). Intense synaptic activity enhances temporal resolution in spinal motoneurons. *PLoS ONE 3*, e3218.

Bhat, B. R. (1961). Some properties of regular Markov chains. *Ann. Math. Statist. 32*, 59–71.

Bibbona, E., P. Lansky, L. Sacerdote, and R. Sirovich (2008). Errors in estimation of the input signal for integrate-and-fire neuronal models. *Phys. Rev. E 78*(1), Art. No. 031916.

Bibbona, E., P. Lansky, and R. Sirovich (2010). Estimating input parameters from intracellular recordings in the Feller neuronal model. *Phys. Rev. E 81*, Art. No. 031916.

Bibbona, E. and A. Rubba (2011). Boundary crossing random walks, clinical trials and multinomial sequential estimation. Preprint, arXiv:1101.4038.

Buonocore, A., A. G. Nobile, and L. M. Ricciardi (1987). A new integral equation for the evaluation of first-passage-time probability densities. *Adv. in Appl. Probab. 19*(4), 784–800.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica 53*(2), 385–407.

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton Mathematical Series, vol. 9. Princeton, N. J.: Princeton University Press.

Feller, W. (1951). Two singular diffusion problems. *Ann. of Math. (2) 54*, 173–182.

Ferebee, B. (1983). An unbiased estimator for the drift of a stopped Wiener process. *J. Appl. Probab. 20*(1), 94–102.

Ghosh, M., N. Mukhopadhyay, and P. K. Sen (1997). *Sequential estimation*. Wiley Series in Probability and Statistics: Probability and Statistics. New York: John Wiley & Sons Inc. A Wiley-Interscience Publication.

Giorno, V., P. Lánský, A. G. Nobile, and L. M. Ricciardi (1988). Diffusion approximation and first-passage-time problem for a model neuron. III. A birth-and-death process approach. *Biol. Cybernet. 58*(6), 387–404.

Giorno, V., A. G. Nobile, L. M. Ricciardi, and S. Sato (1989). On the evaluation of first-passage-time probability densities via nonsingular integral equations. *Adv. in Appl. Probab. 21*(1), 20–36.

Giraudo, M. T., P. Greenwood, and L. Sacerdote (2011). How sample paths of leaky integrate and fire models are influenced by the presence of a firing threshold. *Neural Computation 23*(7), 1743–1767.

Giraudo, M. T. and L. Sacerdote (1999). An improved technique for the simulation of first passage times for diffusion processes. *Comm. Statist. Simulation Comput. 28*(4), 1135–1163.

Girshick, M. A., F. Mosteller, and L. J. Savage (1946). Unbiased estimates for certain binomial sampling problems with applications. *Ann. Math. Statistics 17*, 13–23.

Höpfner, R. (2007). On a set of data for the membrane potential in a neuron. *Math. Biosci. 207*(2), 275–301.

Iacus, S. M. (2008). *Simulation and inference for stochastic differential equations*. Springer Series in Statistics. New York: Springer. With R examples.

Jahn, P., R. W. Berg, J. Hounsgaard, and S. Ditlevsen (2011). Motoneuron membrane potentials follow a time inhomogeneous jump diffusion process. *J. Comput. Neurosci.*, To appear, available online.

Jung, S. and K. Kim (2004). On the estimation of the binomial probability in multistage clinical trials. *Statistics in Medicine 23*(6), 881–896.

Lansky, P., L. Sacerdote, and F. Tomassetti (1995). On a comparison of Feller and Ornstein-Uhlenbeck models for neural activity. *Biol. Cybern. 73*(5), 457–465.

Lansky, P., P. Sanda, and J. He (2006). The parameters of the stochastic leaky integrate-and-fire neuronal model. *J. Comput. Neurosci. 21*(2), 211–223.

Lansky, P., P. Sanda, and J. He (2010). Effect of stimulation on the input parameters of stochastic leaky integrate-and-fire neuronal model. *J. Phys-Paris 104*(3-4), 160–166.

Liptser, R. S. and A. N. Shiryaev (2001). *Statistics of random processes. II* (expanded ed.), Volume 6 of *Applications of Mathematics (New York)*. Berlin: Springer-Verlag. Applications, Translated from the 1974 Russian original by A. B. Aries, Stochastic Modelling and Applied Probability.

Meyn, S. and R. L. Tweedie (2009). *Markov chains and stochastic stability* (2nd ed.). Cambridge: Cambridge University Press.

Novikov, A. A. (1972). Sequential estimation of the parameters of processes of diffusion type. *Mathematical Notes 12*, 812–818.

Paninski, L. (2006a). The most likely voltage path and large deviations approximations for integrate-and-fire neurons. *J. Comput. Neurosci. 21*, 71–87.

Paninski, L. (2006b). The spike-triggered average of the integrate-and-fire cell driven by Gaussian white noise. *Neural Comput. 18*, 2592–2616.

Picchini, U., S. Ditlevsen, A. De Gaetano, and P. Lansky (2008). Parameters of the diffusion leaky integrate-and-fire neuronal model for a slowly fluctuating signal. *Neural Computation 20*, 2696–2714.

Ricciardi, L. and S. Sato (1988). First-passage-time density and moments of the Ornstein-Uhlenbeck process. *J. Appl. Prob. 25*, 43–57.

Różański, R. (1989). Markov stopping sets and stochastic integrals. Application in sequential estimation for a random diffusion field. *Stochastic Process. Appl. 32*(2), 237–251.

Sacerdote, L. and M. T. Giraudo (2012). Stochastic integrate and fire models: a review on mathematical methods and their applications. In Bachar, Batzel, and Ditlevsen (Eds.), *Stochastic Biomathematical Models with Applications to the Insulin-Glucose System and Neuronal Modeling*. Springer.

Sacerdote, L. and F. Tomassetti (1996). On evaluations and asymptotic approximations of first-passage-time probabilities. *Adv. in Appl. Probab. 28*(1), 270–284.

Siegert, A. J. F. (1951). On the first passage time probability problem. *Physical Rev. (2) 81*, 617–623.

Sørensen, M. (2011). Estimating functions for diffusion-type processes. In M. Kessler, A. Lindner, and M. Sørensen (Eds.), *Statistical Methods for Stochastic Differential Equations*. Chapmann and Hall.

Sørensen, M. (1983). On maximum likelihood estimation in randomly stopped diffusion-type processes. *Internat. Statist. Rev. 51*(1), 93–110.

Sørensen, M. (1999). On asymptotics of estimating functions. *Braz. J. Probab. Stat. 13*(2), 111–136.

Tang, C. Y. and S. X. Chen (2009). Parameter estimation and bias correction for diffusion processes. *J. Econometrics 149*(1), 65–81.

Whitehead, J. (1986). On the bias of maximum likelihood estimation following a sequential test. *Biometrika 73*(3), 573–581.

Table 6: Comparison between theoretical mean first passage step and sample averages.

	CASE 1	CASE 2	CASE 3	CASE 4
$\mathbb{E}(N)$	33.83	33.83	100.50	98.99
sample average	33.72	34.28	100.73	101.61

A Numerical details

In this Section we provide some details about the numerical procedure used in the simulation studies, both to simulate the sample and to compute the estimated values. We worked in the R environment. Many of our routines have been designed modifying and adapting functions that were implemented in the `sde` package which is thoroughly documented in Iacus (2008).

A.1 Simulations

The simulation of diffusion processes up to their first-passage time through a barrier b requires some care. If at a given time the process was at level $x_n < b$, and we generate the next point and find $x_{n+1} < b$ we cannot assure that the underlying continuous process did not cross the barrier between the two points. If we stop the simulation only when $x_{n+1} \geq b$, we significantly overestimate the first-passage time. To solve this problem two competitive methods were proposed (Giraudo and Sacerdote, 1999; Baldi and Caramellino, 2002). For each couple of simulated points x_n and x_{n+1} (if both are below b), the probability p of the process crossing the threshold between the two points is evaluated and a corresponding Bernoulli random variable is generated: if you get 1 a crossing occurred and x_n is the last point of the path, while 0 means no crossing and the simulation continues. The first method is slightly more accurate when the discretization step gets larger, the second much faster to compute. We choose the second. To avoid influence of the use of the same approximation both in the simulation scheme and in the estimation method, we simulated with a smaller discretization step w.r.t. the one considered for estimation. To assess the accuracy of the simulation we compare the mean first-passage time estimated from the simulations with the one prescribed by the theory. In particular, for the WD we calculate analytically the probability that the passage occurs between two steps of the discretization and the quantity

$$\mathbb{E}(N) = \sum_{n=1}^{\infty} n\Delta \cdot \mathbb{P}((n-1)\Delta < T \leq n\Delta),$$

which is a discretized version of the mean first-passage time. A comparison between $\mathbb{E}(N)$ and its sample values derived from the simulations shows good agreement as reported in Table 6, the second row is already reported in Table 1.

A.2 Different implementation of formula (10)

Another approximation of the distribution of the first-crossing time is the following,

$$G_\theta^b(\Delta|x) = 1 - \int_{-\infty}^{S} f_\theta^b(X_\Delta = y \,|\, X_0 = x)\, dy. \tag{21}$$

In most cases the numerical evaluation of this integral is much slower than using (10) without providing better performance. Nevertheless, there might be occasions where this alternative turns out to be useful. In particular, the possibility of calculating one of the integrals in (10) or (21) analytically would drastically speed up the algorithm.

In particular, for the OU process we can approximate $f_\theta^b(y, \Delta|x)$ in (21) by expression (12) for the Wiener process, which is the first order approximation in Δ of (15), and we can calculate the integral analytically. If this expression replaces (10), the algorithm becomes significantly faster but less precise, especially if the discretization step is not extremely small. Numerical evaluation of (10) in the parameter settings used here is in any case reasonably fast so we suggest its adoption.

A.3 Minimization algorithm

To minimize the negative log-likelihood function we used the standard Nelder-Mead algorithm provided by the R function `optim`. There are restrictions on the admissible values for some parameters: In the SR model $\mu \leq \sigma^2/2$, and σ and β have to be positive in the SR and in the OU model. The likelihood function is evaluated as `NA` (missing value) if the minimizer tries to calculate it for parameters out of this range, and the minimum is calculated just among the admissible values. The effect of this constraint can be seen in Figures 1 and 4 both in the densities and in the Q-Q plots relative to estimation from a single trajectory. As soon as the estimates get better the effect is lost. The box-constrained algorithm, which is denoted "L-BFGS-B" in R, turned out not to be feasible as it halts if the likelihood function returns `NA`s of infinite values. Especially for the SR model the transition density (even in absence of a barrier) might be problematic to evaluate when the parameters provided by the minimizer are not close to the true ones. In this case we need to be sure that the function returns `NA` when it is not evaluated with satisfactory precision. For the transition density of the SR model the R functions `dchisq` and `pchisq` turned out to be the best choice among the different possibilities discussed in (Iacus, 2008, Section 3.1.3). Nealder-Mead minimizers require reasonable initial values. These values are provided by estimators that can be calculated explicitly. We used the standard choices suggested in the literature in absence of a barrier. The initial estimators for the WD model are

$$\hat{\mu} = \frac{X_N}{(N-1)\,\Delta}; \qquad \hat{\sigma}^2 = \frac{\sum_{i=1}^{N} (X_i - X_{i-1} - \hat{\mu}\,\Delta)^2}{(N-1)\,\Delta}.$$

The initial estimators for the OU model are

$$\hat{\beta} = -\frac{1}{\Delta}\log\left(\frac{\sum_{i=1}^{N}(X_i - \bar{X})(X_{i-1} - \bar{X})}{\sum_{i=0}^{N}(X_i - \bar{X})^2}\right); \quad \hat{\mu} = \hat{\beta}\bar{X}; \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^{N}(X_i - X_{i-1})^2}{(N-1)\,\Delta}.$$

The initial estimators for the SR model are

$$\hat{\beta} = -\frac{1}{\Delta}\log\left(\frac{N\sum_{i=1}^{N}\frac{X_i}{X_{i-1}} - \sum_{i=1}^{N}X_i\sum_{i=1}^{N}\frac{1}{X_{i-1}}}{N^2 - \sum_{i=1}^{N}X_{i-1}\sum_{i=1}^{N}\frac{1}{X_{i-1}}}\right);$$

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N}X_i + \frac{\mathrm{e}^{-\hat{\beta}\Delta}(X_N - x_0)}{N\hat{\beta}(1-\mathrm{e}^{-\hat{\beta}\Delta})};$$

$$\hat{\sigma}^2 = \frac{2\hat{\beta}\sum_{i=1}^{N}\frac{1}{X_{i-1}}\left(X_i - \mathrm{e}^{-\hat{\beta}\Delta}X_{i-1} - \frac{\hat{\mu}}{\hat{\beta}}(1-\mathrm{e}^{-\hat{\beta}\Delta})\right)^2}{(1-\mathrm{e}^{-\hat{\beta}\Delta})\sum_{i=1}^{N}\frac{1}{X_{i-1}}\left[\frac{\mu}{\beta}(1-\mathrm{e}^{-\hat{\beta}\Delta}) + 2\mathrm{e}^{-\hat{\beta}\Delta}\,X_{i-1}\right]}.$$